

A COMPREHENSIVE ANALYSIS OF SOCIAL NETWORK MINING

DIPESH P. JOSHI & KAMAL K. SUTARIA

Department of Computer Engineering, VVP Engineering College, GTU, Gujarat, India

ABSTRACT

Social network analysis is a new research field in data mining. Social network analysis has gained significant attention in recent years, largely due to the success of online social networking and media-sharing sites, and the consequent availability of a wealth of social network data. A social network can be viewed as a complex interconnection of social entities. Mining a community is the task of grouping these social entities together on the basis of their linked pattern. A lot of research has been done on this subject but most of them were only concerned with basic clustering algorithm and graph mining. There are many problems regarding social network analysis such as clustering, community detection, graph creation, link prediction. The clustering in social network analysis is different from traditional clustering.

KEYWORDS: SNAM, SNA, Data Mining, Clustering Algorithm, Link Prediction

INTRODUCTION

Social network analysis emerged as an important research topic in sociology. Most of the early works were conducted on data collected from individual's in particular social settings, in order to study specific social phenomena. The analysis was usually carried out as a "field study" on small communities, gathering data through questionnaires, interviews, and other labor-intensive methods. Comprehensive study on SNA focus on structural component of social network, methods for social network mining, issue regarding social network mining and tools used for social network mining.

ISSUE IN SOCIAL NETWORK ANALYSIS AND MINING

Data Acquisition and Preparation

In the early days of social network analysis, the biggest hurdle was collection of relevant data. There were no "automatic" methods to collect data and, as in most of social science research, data collection was done by performing interviews and often small-scale group studies with volunteers. Now a days, the collection of raw data available from online sources (e.g., Web) and offline sources (e.g., call data) and is much easier, and while data quality has always been an important issue and there are new challenges specific to social networks that include the computational complexity in analysing networks of millions or billions of nodes and the integration of multiple data sources in treating connections.^{[11][15]}

Explicit and Implicit Connections

Explicit connection can be discovered by explicitly their "friends" or connections, "join" a group, "follow" a user, and accept a "friendship" request. Implicit connections can be discovered from user's activities by analyzing extensive and repeated interactions between users.^[11]

Community Detection

The most well known structural problem in the context of social networks is that of community detection. A social community can be formed on web by the people sharing hobbies, working together, living together or having similar ideas

about a subject. Community detection is a technique to classify the network nodes in a group or in a community within which the attribute of these nodes is maximized.^{[2][4][8][9]}

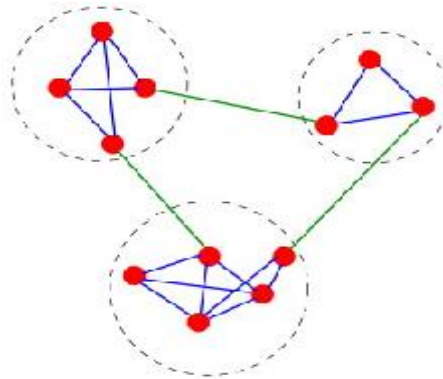


Figure 1: A Simple Graph with Three Communities^[12]

Anonymization

The graph of social connections of users can be a rich source of information and may be used to discover personal information about users. The objective of protecting the privacy of individuals represented in databases means finding the right path between data hiding and data disclosure. A basic operation in data Anonymization is to perturb the data so that individual values are hidden, while still being able to recover useful information, such as the distribution of the data values or rules and patterns in the data.^{[11][15]}

METHODS OF SOCIAL NETWORK ANALYSIS AND MINING

Two main kinds of data that are analysed

- **Linkage-Based and Structural Analysis**

In linkage-based and structural analysis, we construct an analysis of the linkage behaviour of the network in order to determine important nodes, communities, links, and evolving regions of the network.

- **Adding Content-Based Analysis**

Many social networks contain a tremendous amount of content which can be leveraged in order to improve the quality of the analysis. For example, a photograph sharing site contains a tremendous amount of text and image information in the form of user-tags and images. Similarly, blog networks, email networks and message boards contain text content which is linked to one another.

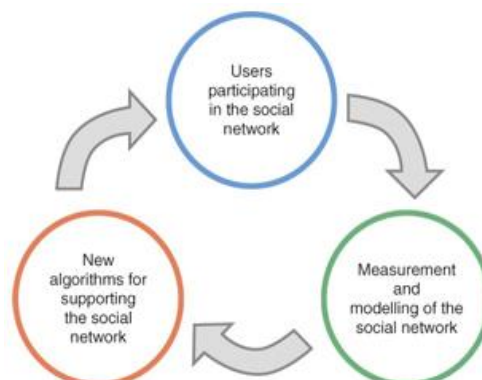


Figure 2: Life Cycle of Social Network Mining.^[24]

There are many techniques and method available for social network mining.

Hierarchical Algorithms

It gives hierarchical decomposition of the nodes of the social network. Such hierarchical methods have been used traditionally in sociology. A property of these methods is that they return not just a flat partitioning of the network into communities, but a hierarchy of communities and sub communities.^[12]

Modularity Maximization

Girvan and Newman [2002] proposed a measure of evaluating the quality of a partitioning of a network into communities, and selecting the best community partitioning from a hierarchal decomposition. The measure is called modularity, and is defined as the fraction of edges that fall within communities minus the same fraction if edges were assigned at random.^[12]

Graph-Partitioning Algorithms

The fundamental problem that is trying to solve is that of splitting a large irregular graphs into k parts. The partitioning is usually done so that it satisfies certain constraints and optimizes certain objectives. The most common constraint is that of producing equal-size partitions, whereas the most common objective is that of minimizing the number of cut edges.^[12]

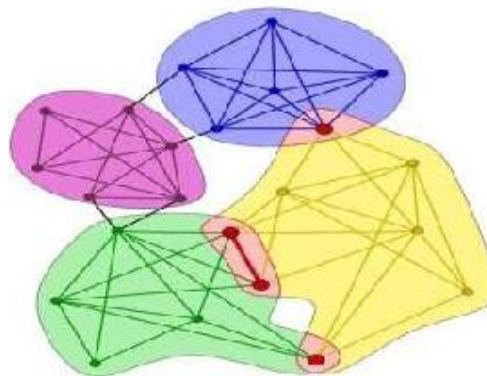


Figure 3: A Simple Graph with Communities

TOOLS USED FOR SOCIAL NETWORK ANALYSIS AND MINING(SNAM)

There are many tools used to implement social network mining. Some tools given below that are used to analyse it:

Gephi

Gephi is an interactive visualization and exploration platform for all kinds of networks and complex systems, dynamic and hierarchical graphs. It is a tool for people that have to explore and understand graphs. The user interacts with the representation; manipulate the structures, shapes and colors to reveal hidden properties. It uses a 3D render engine to display large networks in real-time and to speed up the exploration. A flexible and multi-task architecture brings new possibilities to work with complex data sets and produce valuable visual results.^{[18][21][19]}

Graphviz

Graphviz is open source graph visualization framework. It has several main graph layout programs suitable for social network visualization.^{[19][22]}

Sondy

SONDY is a tool for analysis of trends and dynamics in online social network data. SONDY helps end-users like media analysts or journalists understand social network users interests and activity by providing emerging topics and events detection as well as network analysis functionalities.^{[16][20]}

Neo4j

Neo4j is a graph database. It is an embedded, disk-based, fully transactional Java persistence engine that stores data structured in graphs rather than in tables.^{[19][23]}

Many other tools are available for graph mining as well as community detection and topic detection.

CONCLUSIONS

This paper presents analytical study and current trend in social network mining .here we give basic understanding of social network and data mining. In order to collect data and process data which step to be conducted and which basic algorithm are available to implement. There are many tool, software and framework available for social network analysis .By using Sondy, we can implement our algorithms well as dataset to analyse current trend.Gephi is best tool with rich plug-in set to analyse social network.

ACKNOWLEDGEMENTS

We thank our institute for their support in research work progress. We greatly appreciate the reviewers and editor for their valuable suggestions and comments to improve this work.

REFERENCES

1. S Wil M.P. van der Aalst and Minseok Song (2004), Mining Social Networks: Uncovering Interaction Patterns in Business Processes, springer
2. Vincenzo Nicosia , Dipartimento di Ingegneria Informatica e delle Telecomunicazioni Università di Catania – Italy, lecture notes on Modularity for community detection: history,perspectives and open issues
3. Wangqun Lin, Xiangnan Kong, Philip S. Yu, Quanyuan Wu, Yan Jia, Chuan Li, Community Detection in Incomplete Information Networks,Presented in WWW 2011
4. Sadi, Sercan, Şima Etaner-Uyar, and Şule Gündüz-Öğüdücü.(2009) "Community detection using ant colony optimization techniques." *Proc. Int. Conf. Soft Computing (MENDEL'09)*.
5. Creamer, G., & Stolfo, S. (2009). A link mining algorithm for earnings forecast and trading. *Data mining and knowledge discovery*, 18(3), 419-445.
6. Yiannis Kompatsiaris,(2013), Social Networks Mining for Innovative Applications and Users Well-Being, FP7 ICT Work Programme 2013 Consultation Networked Media
7. Guy, I., Avraham, U., Carmel, D., Ur, S., Jacovi, M., & Ronen, I. (2013, May). Mining expertise and interests from social media. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 515-526). International World Wide Web Conferences Steering Committee.
8. Griechisch, Erika, and András Pluhár.(2011). "Community Detection by using the Extended Modularity." *Acta Cybern.* 20.1 (2011): 69-85.

9. Kai-Yang Chiang, Department of Computer Science, University of Texas at Austin, Lecture notes on community detection
10. Leskovec, Jure, Kevin J. Lang, and Michael Mahoney.(2010). "Empirical comparison of algorithms for network community detection." *Proceedings of the 19th international conference on World wide web*. ACM, 2010.
11. Bonchi, Francesco, et al.(2011) "Social network analysis and mining for business applications." *ACM Transactions on Intelligent Systems and Technology (TIST)*2.3 (2011): 22.
12. Fortunato, Santo.(2010) "Community detection in graphs." *Physics Reports* 486.3 (2010): 75-174.
13. Blondel, Vincent D., et al. (2008). "Fast unfolding of communities in large networks. " *Journal of Statistical Mechanics: Theory and Experiment* 2008.10 (2008): P10008.
14. Ellison, Nicole B.(2007). "Social network sites: Definition, history, and scholarship. " *Journal of Computer-Mediated Communication* 13.1 (2007): 210-230.
15. Lin, C. Y., Wu, L., Wen, Z., Tong, H., Griffiths-Fisher, V., Shi, L., & Lubensky, D. (2012). Social network analysis in enterprise. *Proceedings of the IEEE*,100(9), 2759-2776.
16. Guille, Adrien, et al. (2013). "Sondy: An open source platform for social dynamics mining and analysis." *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*. 2013.
17. Fire, Michael, Rami Puzis, and Yuval Elovici.(2013). "Organization Mining Using Online Social Networks." *arXiv preprint arXiv:1303.3741* (2013).
18. Bastian M., Heymann S., Jacomy M. (2009). *Gephi: an open source software for exploring and manipulating networks*. International AAAI Conference on Weblogs and Social Media.
19. http://en.wikipedia.org/wiki/Social_network_analysis_software
20. <http://mediamining.univ-lyon2.fr/people/guille/software.php>
21. <https://gephi.org/>
22. www.graphviz.org/
23. www.neo4j.org/
24. Albert Ching-man Au Yeung and Tomoharu Iwata, Research on Social Network Mining and Its Future Development, NTT Technical Review

